SURP Project Notes

Week 1

- Dark energy drives the acceleration of the universe
- Things close together are more highly correlated
- Comoving separation
- Use Ly-A to compute the correlation function and find the location of the bump
- Bumps can be caused by metal absorption
- High column density absorbers (main source of interference we will analyze)
- Comoving coordinate system is what would be seen by an observer moving with the expanding universe over time
- Universe from the start was like a bag of pebbles dropped in a pond, creating acoustic ripples in the cosmic plasma → these sound waves are seen as acoustic oscillations in maps of the CMB
- "'Trees' in the forest are the highly photoionized sheets, filaments, and halos that result from cosmic structure formation in a Universe with gravitationally dominant cold dark matter and with an approximately uniform ionizing background"
- BAO has anticipated constraints on dark energy
- Areas that might be constrained: inflation, neutrino masses, "coldness" of dark matter (i.e. that the particles move slowly compared to the speed of light)
- Sv = survey validation from test period
- Weights based on IVAR (inverse variance)
- OHIO mocks by Naim which we won't really use → probably London mocks instead for us
- Mocks = synthetic data sets
- Photon jumping from 1st to 2nd energy level shell emits LyA
- Emission spectra
    - Strong emission line ~121 nm is LyA
    - Lyman limit is min energy for photon at first shell to escape hydrogen atom (everything below limit absorbed)
    - Lyman alpha is max absorption
- Working w/ damped systems (known as DLAs)
    - When light is close to galaxy clusters
    - Circumgalactic medium
    - Complete drop to 0 w/ long wings instead of small absorption dip, wings have their own dips (main contaminant!/indicates location of galaxy)
    - Want impact of wings on correlation
- Quasar emits spectrum and travels through universe to us → light redshifted from expansion (doppler effect)
    - Produces dip at original 121nm
    - Absorption when light hits gas cloud
    - Always neutral hydrogen
- General relativity

    - Think about changing geometry
        - Very nontrivial distance measurements (curvature, expansion, etc.)
        - Integrate over effects
    - Dark matter bends spacetime too
    - Acceleration of expansion means there is some repulsive force
    - DESI will be able to help measure the cosmological constant
    - Want to evaluate if dark energy changes with time
    - Need to account for collective gravity of weakly interacting particles like neutrinos (not massless)
- Correlation function
    - Fourier transforms → two-point statistic and power spectrum (need to know both!)
    - BOSS/eBOSS survey from SDSS
    - Data is a function of two parameters across and along line of sight
    - Cross correlate pixels of spectra and assign to bins → repeat for entire matrix to produce data
    - Compare to theory which comes from model power spectrum (fourier transform)
    - Data not isotropic, one reason bc we are in redshift space
    - RSD = redshift space distortions → 2D function
    - Code will work on model and compare to mock data w/ one main contaminant
- Power spectrum
    - Real space to Fourier space
    - Small bump ~100 Mpc/h → increased chance of finding two objects (BAO effect)
        - h is H0/100 for hubble constant H0
        - k is the inverse (h/Mpc) → small k means large scales
    - Power law of k w/ damped oscillations
    - Epoch of matter radiation equality
- Vega package reads P(k) function (power spectrum of matter/distribution of all matter in the universe → transfer to just LyA in 2D → Fourier transform to correlation)
    - Boltzmann eqs. and GR
    - Compare theory to correlation data
    - Multiply power spectrum template by bias (can break down at small scale, ignore this)
        - Kaiser/linear theory model
- High-density matter areas (density peaks only which creates bias) will condense into galaxies but low-density will exist in clouds
- Redshift space distortions = influence of peculiar velocities
- Decompose power spectrum of LyA into spherical harmonics and then Fourier transform
- Don’t compare pixels in the same forest
- 1D: neutrino masses, dark matter
- 3D: properties of dark energy, curvature of universe, initial conditions of universe

- Take quasar light out of the equation to analyze absorption alone
- Delta flux equation (overdensity) = find flux per pixel / mean absorption
- HCD = high column density absorber
    - HCD > 10^{17} N(HI)/cm^2
        - Aka sub-DLA
    - Includes all DLAs
        - DLA > $10^{20.3}$ N(HI)/cm^2
        - Trace cosmology
        - What if we don't cut them out? Influence of bottoming out the absorption dips → improves error bars (group in France w/ different mocks)
    - Regular LyA line < 10^{17}
    - Density inside a column (length of line of sight vs density in length)
- Power spectrum transformation
    - $P(k,uk) = b^2 (1+Buk^2)^2 P(k)$
    - b = bias factor
    - B = redshift distortion parameter
- LLS = Lyman limit systems
- Absorption profiles from HCDs affect measurements of b and B
- Anisotropic = having a physical property that has a different value when measured in different directions
- Measure contamination of quasar spectra by the damping wings of HCD absorbing regions of neutral hydrogen on correlations in the 3D LyA forest
    - Accomplished by measuring 3D flux power spectra from a cosmological hydrodynamical simulation as a function of column density of HCD contamination and redshift

<u>Week 2</u>

- Fiducial = reference
    - For fibers on DESI
    - Or cosmology from Planck 2018 model
- Visual aberrations
    - Astigmatism: elongated
    - Coma: brighter on one side
- 8 square degree fiber area
- DESI focal plane ~ 6 moons in diameter
- 3 spectrograph channels: red, blue, near-infrared
- CCD = charge-coupled device which measures photons —> stored electrons
- "In theory, theory and practice are the same, but in practice they are not"
- Dip in collimator reflection (absorption feature in spectrum) comes from imperfection in coating of collimator mirrors
- 13.6 ev to ionize hydrogen

- Transition lines order: alpha(2→1, ~1216A), beta(3→1), gamma(4→1), delta(5→1), … (infinity → 1, ~912A)
- Radiative transfer
    - Emission negligible compared to absorption
- Equivalent width
    - Area of absorption feature dip equal to integral rectangle
- Can figure out optical depth but hard to determine how deep in a gas cloud the photons are absorbed
- Milky Way galaxy is a DLA
- Use cgs units
- LLS = Lyman limit systems (10^{17-19} cm^{-2})
- Universe filled with low density gas between galaxies = intergalactic medium/IGM → mostly hydrogen and mostly ionized as opposed to neutral
    - Ionized bc of abundance of UV photons
    - IGM in photoionization equilibrium
- Cosmic UV background caused by star formation/AGNs
- Relationship between temperature and density in the IGM
- Quasars surrounded by very high density gas which produces their emission lines (Lyman peaks)
- Bayesian statistics
    - Belief in a certain event happening
    - Base theorem = P(x | y) = P(y | x) P(y) / (P(y)
    - X: model parameter vector theta, Y: data → P(theta | D, M) = P(D | theta, M) P(theta | M) / P(D | M)
    - P(D | theta, M) = likelihood L
        - L proportional to $e^{((-\frac{1}{2})(\chi^2))}$
    - P(theta | M) = pi = prior
        - Chosen value from best estimates
        - Top hat model (wide rectangular prior)
    - P(theta | D, M) = posterior
        - 4 dimensional in our case (bias, beta, ap, at)
    - Need full shape of fit if not a Gaussian fit to determine posterior
    - P(D | M) = Bayesian evidence
        - Normalization of posterior
        - Integral of likelihood x prior x dtheta
    - Often used in machine learning

<u>Week 3</u>

- Need to code in the same space, i.e. Fourier space
- Doppler and pressure broadening → Gaussian and Lorentzian combination → Voigt profile

- Faddeeva function in Scipy is the real part of Voigt
- About 10% of forests have DLAs and of those about 10% of that forest is influenced by the DLA
- Correlate pixels between comparison forests
- More column density → more absorption (see eq. in bb 2012 paper)
- Want to model wings
- Will have Voigt profile of column density and centering variable, V(N,v'-v)
- Integrate over all possible column densities and then over the line of sight which would be very hard, DO NOT WANT TO DO THIS so do it in Fourier space instead
- Fourier transform taken of just the box around Voigt profile in the past, ignore wings and approximates exponential/Fourier transform of Lorentzian
- Manipulate mocks (theoretical modeling for simulated spectra)
- Will need to take Voigt profile code and put it into code in vega
- Evaluate mock cases for w and w/out noise, distortion matrix DM, and different column densities
- Frequency intervals where natural broadening > Doppler broadening are the damping wings of the profile function
- Frequency distribution of HI column densities
    - Increases w decreasing NHI
- Prochaska, Herbert-Fort, and Wolfe 2005 paper on the frequency distribution for all damped LyA systems identified by SDSS DR3

<u>Week 4</u>

- Colore code
    - Produces transmission files
    - Density profiles throughout the universe
    - No quasars, no physics, just a box with cosmology
    - Creates DLA catalog from when skewers go over highly dense regions
    - QQ (quick quasars) → simulated DESI spectra → true continuum
    - Noise level, distortion matrix, and column density variables
- Power spectrum of auto and cross correlations
    - Difference will be seen between them if the bias HCD parameter measured by the Voigt profile/Fourier transform is wrong
    - P/Plim vs. k plot
        - NHI values will change the length of decay
        - Larger values = shorter extension before decay
        - Straight line across line of sight
        - DNE if exactly along line of sight
- Terminal instructions
    - Pip install picca (to install package)
    - Picca_ tab (to examine package options)

    - Picca_compute_fvoigt.py -h (to call help documentation)
    - Ls .conda/ (view options)
    - Ls .conda/envs/
    - Ls .conda/envs/vega/
    - Ls .conda/envs/vega/bin/ (choose y to display all options)
    - Cd .conda/envs/vega/bin/ (move to desired directory/location)
    - Pwd (print working directory/create path name)
- CDDF: f(NHI)~NHI^-1 for NHI<10^21 cm^-2 and f(NHI)~NHI^-3 for higher column densities

<u>Week 5</u>

- Smaller DLAs = smaller wings = effect on smaller scales (shifts right, probability = 1 for longer)
- Seeing edges is bad → need more bins

<u>Week 6</u>

- Bias scales the correlation up and down
- Current file
    - All HCD → NHI = 20 (COULD BE THE CAUSE OF WEIRD MODEL SHAPE)
    - Noiseless
    - True continuum (could be a bug in the analysis, it is a new area of active development)
- New file
    - HCD → All NHI range
    - Noisy
    - Not true continuum/normal continuum
    - Realistic universal conditions
- Another file
    - No HCD
    - Noiseless
    - True continuum
    - Compare to current file to determine if error comes from HCDs or a bug
- Alpha, phi = BAO parameters
- Alpha s, phi s, = nuisance AP parameters
- Goal to understand the impact of HCDs on these four parameters
- Fix these = 1 as an initial “gold standard” for cosmology (as measured in the CMB/Planck)
- LyA working group goals
    - Use simulations to measure cosmological constants
    - Use mocks to recover parameters without bias
    - Measure with data
- Fits

        - Alpha, phi = 1
            - How good is the fit?
        - Alpha s, phi s = 1; fit alpha, phi (address LyA goal w only HCDs)
        - Fit all four
- Use chi-squared
    - Minimum chi-squared finds the best fit
    - Want reduced chi-squared of ~1
        - Overfit
            - Data points <= parameters
            - I.e. degrees of freedom <= 0
            - Reduced chi-squared less than 1
        - Underfit
            - Obviously wrong
            - Data points >> parameters
            - Reduced chi-squared more than 1
    - FCN (function) = chi-squared
- When we add HCD, there will be two extra parameters (bias hcd and beta hcd)
- Minimizing chi-squared maximizes the likelihood function
- Underfit considered bad
- When reduced chi-squared = 1 it is sometimes referred to as reaching the noise level
- Degrees of freedom may be inconsistent for nonlinear models
- Chi-squared usually has a large uncertainty
- Weird graph shape results do not appear to be the result of a bug → influence from HCDs → high abundance of DLAs in NHI=20
- Fiducial cosmology is taken as real for the mocks
- Alpha s inconsistency
    - Nuisance parameter
    - Does not tell us any cosmological information
    - Isotropic scale/amplitude of power spectrum
    - Multiple model/data unknowns
- Combination of auto/cross-correlations helps correct parameter uncertainty
- We care about parameter offset from 1
- These data points are highly correlated → noise
    - Does not discredit chi-squared analysis bc of covariance matrix consideration
- True continuum model is still a work in progress
- The NHI=20 case does not converge to 0
- Delta = flux / ((mean flux * continuum) - 1)
- We compute the correlation function of delta
- Expect average delta (function of lambda/wavelength) to be 0
- If not 0 —> mean flux is slightly wrong
- Effect is worse for greater HCDs

- Astropy fits reader is slow for large number of files —> use fitsio written in C
- Data on a sphere, healpixels
- 1 forest per hdu —> hdul list of all
- Counts = sum of weights
- Simplemock is good, decent correlations
- Baselinemock is failing —> delta error bc model assumes 0
- Covariant matrix gives error on data —> error large and correlated

Week 7
- Regions above 0 are underdense and below 0 are overdense
- Reminder! Python starts counting from 0
- Want to get delta/trans close to 0/1

Week 8
- Deltas are important results
- Problems w the true continuum fits
- Cooked mocks have distortion
- The bottom right plot in the 4 wedge is along the line of sight

Week 9
- HCDs reduced bias
- Posterior probability
- 5 sigma is the physics standard for new discoveries
- Deltas problem good for lya meeting but too complex for surp
- Lose constraining power but less bias
- Early fall of line in picca correlation is the impact of hcds
- K is k parallel, along the line of sight in picca graph
- Hcds are localized in the universe but spectra impact extends to large regions bc of the wings → dropoff in correlations along line of sight
- Only happens at very large k (don't need to get to this point in explanation)
- K is the wave number in Fourier space
- Machine learning can only detect HCDs above NHI=20
- Picca bump might be coming from width of DLA (current guess)
- Rogers model was a best estimation based off of exponential appearance
- Measure AP effect to constrain cosmology

Week 10
- Cooked models includes redshift distortion
- Corrected models come from updated deltas as a solution to the failed true continuum
- Fvoigt models include HCD modelling based on run fits files
- Fvoigt adds to the cooked models

- Fvoigt does not fully fix the bias bc of the distortion problem
- Bias reduced when the separation is extended from 10 Mpc to 25 Mpc